Is the Natural Twinning Rate Still Declining?

To the Editor:

An unexplained reduction of one third in twinning rates was observed in Western countries from the 1930s through the 1970s[1] followed by a sharp rise from the 1980s onward due mainly to infertility treatment.[2,3] There is concern about a general decline in fecundability, and it is unknown whether the "natural twinning rate" (which is a marker of fecundability[4]) is still declining beneath the increase of twinning due to infertility treatment. We used a follow-up study of more than 60,000 pregnant women, who reported on fertility treatment during pregnancy, to show that the decline in natural twinning rates has apparently stopped.

The Danish National Birth Cohort is a nationwide study that enrolled women early in pregnancy. Approximately 60% of all eligible women were informed about the study, depending on the cooperation of general practitioners, and approximately 60% of the informed women agreed to participate.[5] The women were typically enrolled in the study at the first visit to the general practitioner (usually in the first trimester). At this time, the participants provided a blood sample and were interviewed about health-related issues, including potential fetal risk factors and lifestyle. Around the beginning of the third trimester, the participants again provided a blood sample and were again interviewed; topics included waiting time to pregnancy and infertility treatment. We used the National Birth Registry to identify 61,995 singleton and twin births from the 28th gestational week occurring between 1998 and 2001 of women participating in the Danish National Birth Cohort.

We excluded 49 pregnancies from the analyses because they were either triplets or stillbirths with no information on twinning status. Among the remaining 61,946 pregnancies, 1311 (2.1%) resulted in twins; this percentage is comparable to that (2.0%) for the Danish population in the same time period,[3] suggesting no selection into the cohort depending on whether the pregnancy was a twin pregnancy.

Women taking 6 months or longer to conceive were asked whether they had received infertility treatment; among the 3873 women answering positively to this question, 589 (15.2%) had twins. The remaining 58,073 pregnancies, with no infertility treatment reported, resulted in 722 pairs of twins, which corresponds to a natural twinning rate of 1.24% (95% confidence interval = 1.15–1.33%). As expected,[1–3] the natural twinning rate increased with maternal age (from 1.0% in mothers younger than 25 years of age to 1.5% among mothers age 35 an older), with the most marked increase being for opposite-sexed twins who are all dizygotic.

Twinning rates (Fig. 1) and maternal age distribution for births in Denmark in 1930–2001 were obtained through Statistics Denmark. From 1970 to 1980, when the overall twinning rate was historically low, the mean maternal age was very stable, around 26.6 (within a 0.2-year standard deviation), whereas the mean maternal age in the more recent Danish National Birth Cohort was 29.9 ± 4.3 years. Age standardization of the natural twinning rate in the Danish National Birth Cohort to the maternal age distribution of 1980 changed the estimate of twinning rates only slightly from 1.24% (1.15–1.33%) to 1.15% (1.06–1.23%), which is markedly higher than the rate of 0.97% (0.95–0.99%) in the period 1970–1980.

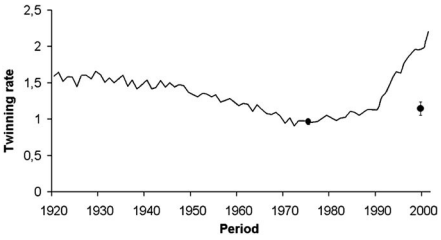

FIGURE 1. The overall twinning rate (per 100 births) in Denmark for 1930–2001. The 2 point estimates are the mean twinning rate for the period 1970–1980 and the twinning rate estimate for women with no fertility treatment in the Danish National Birth Cohort 1998–2001. (This later "natural twinning rate" is standardized to the 1980 maternal age distribution).

This follow-up study provides evidence for the end of the decline in natural twinning and may even suggest an increase in natural twinning rates. The rate of monozygotic twinning is fairly constant over time and place, whereas the frequency of natural dizygotic twinning depends on maternal age, parity, ethnicity, and body mass index.[2,4] However, none of these factors appears to account for the observed, marked changes in natural twinning rates, which remain unexplained.

Anne Maria Herskind
Pediatric Department
Odense University Hospital
Odense, Denmark

Olga Basso
Jørn Olsen
Danish Epidemiology Science Centre
University of Aarhus
Aarhus, Denmark

Axel Skytthe
Kaare Christensen
Danish Twin Registry, Epidemiology
University of Southern Denmark
Odense, Denmark
kchristensen@health.sdu.dk

REFERENCES

1. Rachootin P, Olsen J. Secular changes in the twinning rate in Denmark 1931 to 1977. *Scand J Soc Med*. 1980;8:89–94.
2. Westergaard T, Wohlfahrt J, Aaby P, et al. Population based study of rates of multiple pregnancies in Denmark, 1980–94. *BMJ*. 1997; 314:775–779.
3. Basso O, Nohr EA, Christensen K, et al. Risk of twinning as a function of maternal height and body mass index. *JAMA*. 2004;291:1564–1566.
4. Basso O, Christensen K, Olsen J. Fecundability and twinning. A study within the Danish National Birth Cohort. *Hum Reprod.* 2004;19: 2222–2226.

5. Olsen J, Melbye M, Olsen SF, et al. The Danish National Birth Cohort—its background, structure and aim. *Scand J Public Health*. 2001;29: 300–307.

Mediation Proportion

To the Editor:

Subsequent to the publication of our paper[1] in the January 2005 issue of *Epidemiology*, relevant references[2–4] to the problem of mediation analysis have been brought to our attention from the discipline of drug abuse prevention.

Using the framework of the multivariate normal distribution, MacKinnon[2] explained mediation analysis in the context of prevention and intervention studies with an abundance of examples of problems of mediation. He proposed to measure the intermediate effect by the product of the regression coefficients in the relevant regression equations and denoted the fraction of the intermediate effect to the total effect for the proportion mediated. When the intermediate and response variables are jointly normally distributed, the analysis is straightforward and intuitive, and in this case, MacKinnon's proportion mediated coincides with our proposed mediation proportion, as do the measures of both Freedman et al[5] and Wang and Taylor,[6] as we have pointed out earlier.[1]

In psychology, it is standard to apply structural equations models. Finch et al[3] studied a model of 3 latent variables, each with 3 indicator variables, and through simulations, analyzed the effects of nonnormality of the indicator variables on the estimates of the intermediate effect. The nonnormality was represented by continuous variables with positive skewness and kurtosis. Nonnormal data are common in epidemiology such as when constructing scales on some measured items or indicators. The data may be analyzed (possibly transformed) assuming normality or by grouping into a smaller number of categories, maybe all the way to binary variables. Unfortunately, the statistical conclusions are sensitive to the construction of scales and choice of cut points. Our main motivation for using structural equations models was to provide a fresh approach to defining and estimating the mediation proportion for the discrete or ordered categorical variables commonly met in epidemiology. Such observations may often be embedded in the structural equations framework through threshold models.[1] The threshold model approach is particularly useful when handling ordered categorical data, in which it will often be natural to assume an underlying continuous variable governing the observed data.

MacKinnon et al[4] discussed the concepts of mediation, confounding, and suppression effects in the context of causality.

Susanne Ditlevsen
Niels Keiding
Department of Biostatistics
University of Copenhagen
Copenhagen, Denmark
S.Ditlevsen@pubhealth.ku.dk

Ulla Christensen
Mogens Trab Damsgaard
Department of Social Medicine
Institute of Public Health
University of Copenhagen
Copenhagen, Denmark

John Lynch
Department of Epidemiology
University of Michigan
Ann Arbor, MI

REFERENCES

1. Ditlevsen S, Christensen U, Lynch J, et al. The mediation proportion: a structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. *Epidemiology*. 2005;16: 114–120.
2. MacKinnon DP. Analysis of mediating variables in prevention and intervention research. In: Cazares A, Beatty LA, eds. Scientific methods for prevention intervention research. *NIDA Res Monogr*. 1994;139:127–153.
3. Finch JH, West SG, MacKinnon DP. Effects of sample size and nonnormality on the estimation of mediated effects in latent variables models. *Structural Equation Modeling*. 1997; 4:87–107.
4. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. *Prev Sci*. 2000;1:173–181.
5. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic disease. *Stat Med*. 1992;11:167–178.
6. Wang Y, Taylor JMG. A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics*. 2002;58:803–812.

Case–Crossover Studies

To the Editor:

Sullivan et al recently presented a case–crossover study[1] with interesting null findings, like in other Seattle studies of the same design.[2,3] In contrast to the Boston study of Peters et al,[4] the Seattle associations between ambient concentrations before onset of a myocardial infarction and ambient conditions on control days were substantially smaller. Unfortunately, this is yet another case–crossover study, out of more than 20 to date, that do not either provide or discuss the distribution of the *relevant* exposure term and its implications on statistical power.[5] The exposure term in the case–crossover design is not the daily level of pollutants, but the *difference* between the ambient concentration on the event day and the concentration(s) on some control day(s). We have shown that this difference can be very small for a large fraction of event days, thereby seriously limiting the statistical power to refute the null hypothesis.[5] The relevant distribution of these differences cannot be inferred from the usual tables showing the distribution of the daily levels (Table 2 in Sullivan et al[1]). We believe it is time for a change in reporting of case–crossover studies. Authors and reviewers alike should opt for a summary of the *relevant* exposure term. Otherwise, an important alternative explanation of null findings cannot be evaluated: insufficient statistical power.

Sullivan et al provided a further example for the need to present findings using the *relevant* exposure metric. To

assess the shape of the concentration–response function, they show risk estimates for quintiles of the ambient concentrations. However, a conceptually appropriate presentation would stratify risk estimates by quintiles of the *relevant* exposure term, as defined previously. This might provide different results.

We acknowledge that the sample size was large in this study and that elaboration of the analyses in the appropriate exposure metric may simply confirm the results. We also appreciate the thorough discussion of the potential environmental causes of the null findings in Seattle. However, we believe that case–crossover studies should comply with longstanding traditions of good epidemiologic practice; in particular, they should describe the (design-relevant) exposure distribution and address statistical power. This is particularly important when trying to investigate regional differences in acute effects of air pollution.[2–4]

Nino Künzli
Keck School of Medicine
University of Southern California
Los Angeles, CA
Kuenzli@usc.edu

Christian Schindler
Institute of Social and Preventive Medicine
University Basel
Basel, Switzerland

REFERENCES

1. Sullivan J, Sheppard L, Schreuder A, et al. Relation between short-term fine-particulate matter exposure and onset of myocardial infarction. *Epidemiology*. 2005;16:41–48.
2. Levy D, Sheppard L, Checkoway H, et al. A case–crossover analysis of particulate matter air pollution and out-of-hospital primary cardiac arrest. *Epidemiology*. 2001;12:193–199.
3. Sullivan J, Ishikawa N, Sheppard L, et al. Exposure to ambient fine particulate matter and primary cardiac arrest among persons with and without clinically recognized heart disease. *Am J Epidemiol*. 2003;157:501–509.
4. Peters A, Dockery DW, Muller JE, et al. Increased particulate air pollution and the triggering of myocardial infarction. *Circulation*. 2001;103:2810–2815.
5. Kunzli N, Schindler C. A call for reporting the relevant exposure term in air pollution case-crossover studies. *J Epidemiol Community Health*. 2005;59:527–530.

The authors respond:

Drs. Künzli and Schindler[1] raise an interesting point regarding the relevant exposure description in acute effect air pollution studies. We agree that it is helpful for the data description to include a summary of the exposure variation actually used in the analysis. Similar to time-series analyses, the convention in case–crossover papers has been to report on the unadjusted exposure distribution. This description ignores the fact that the calculation of the health effect estimate relies on a restricted exposure range. For case–crossover studies, the relevant exposure variability is restricted to within-referent windows.

In our study,[2] the square root of the average within-referent window exposure variance was 10.9 $\mu g/m^3$. This quantity ranged from 1.1 to 46.3 $\mu g/m^3$ with an interquartile range (IQR) of 3.8–11.9 $\mu g/m^3$. For time-series studies, the relevant exposure distribution is obtained after the smooth function of time has been removed. The equivalent calculation for case–crossover studies subtracts the mean for each referent window from all observations in that window. This summary yields an IQR of 7.9 $\mu g/m^3$, compared with the unadjusted IQR of 10.6 $\mu g/m^3$ shown in Table 2 of our paper. Although this description of exposure represents a smaller exposure variation than the data description reported in our paper, it does not change our analysis or results.

Our model assumed a linear dose–response. Thus, referent windows with average exposure of 25 $\mu g/m^3$ contribute the same information as referent windows with average exposure of 5 $\mu g/m^3$ provided they have identical exposure variability. We disagree with the suggestion by Künzli and Schindler that our Figure 1 presentation should have been indexed by some other metric than the PM quintile midrange. Because our model is linear, our presentation gives a visual check of the linear dose–response assumption. It also showed a direct comparison of our results with those for the Boston Onset Study.[3]

We also disagree with the assertion of Künzli and Schindler that insufficient statistical power may be an alternative explanation for the null findings in our study. Once study results have been obtained, confidence interval estimates should be used to assess study power.[4,5] Our study's large sample size produced relatively narrow confidence interval estimates. For a 10-$\mu g/m^3$ increase in PM 1-hour before myocardial infarction (MI) onset, we reported 1.05 as the upper limit of our confidence interval estimate. For risk of MI onset in Seattle, this finding can be interpreted as providing evidence against a short-term effect of fine PM larger than a relative risk of 1.05 for a 10-$\mu g/m^3$ increase in PM.

Jeffrey Sullivan
Lianne Sheppard
Astrid Schreuder
Joel Kaufman
Department of Environmental and Occupational Health Sciences
University of Washington
Seattle, WA
sulljh@u.washington.edu

REFERENCES

1. Künzli N, Schindler C. Case–crossover studies [Letter]. *Epidemiology* 2005;16:592–593.
2. Sullivan J, Sheppard L, Schreuder A, et al. Relation between short-term fine-particulate matter exposure and onset of myocardial infarction. *Epidemiology*. 2005;16:41–48.
3. Peters A, Dockery DW, Muller JE, et al. Increased particulate air pollution and the triggering of myocardial infarction. *Circulation*. 2001;103:2810–2815.
4. Rothman KJ, Greenland S. eds. *Modern Epidemiology,* 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998;193.
5. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results [published correction appears in *Ann Intern Med.* 1995;122:478]. *Ann Intern Med*. 1994;121:200–206.

Population Risk Measures

To the Editor:

In my recent commentary,[1] I had noted some confusion over the nomenclature for measures of the potential impact of an intervention on risk in a community. I underestimated the complete range of terminology, however. Soon after my commentary appeared, Noel Weiss pointed out to me that a measure Tom Koepsell and he had called *attributable risk to the population* (*PAR*) in their 2003 textbook[2] is identical to *attributable community risk* (*ACR*), as used in MacMahon et al[3] and my commentary.[1] Koepsell and Weiss refer to the measure commonly called *PAR* as *attributable risk to the population percent* (*PAR%*) and clearly indicate the distinction between the questions addressed by the 2 measures.[2] The percent in "PAR%" refers to the percentage of cases attributable to the exposure, not the percentage of the population who develop the disease due to the exposure.

I also underestimated the complete range of confusion. Although Professor Weiss taught me PAR and PAR% in class in 1977, I did not appreciate the importance of the distinction for a quarter century.

Sholom Wacholder
Division of Cancer Epidemiology and Genetics
National Cancer Institute
Bethesda, MD
Wacholder@NIH.gov

REFERENCES

1. Wacholder S. The impact of a prevention effort on the community [Commentary]. *Epidemiology*. 2005;16:1–3.
2. Koepsell TD, Weiss NS. *Epidemiologic Methods: Studying the Occurrence of Illness*. New York: Oxford University Press; 2003;196–203.
3. MacMahon B, Pugh TF, Ipsen J. *Epidemiologic Methods*. Boston: Little Brown; 1960.